DOCUMENT RESUME

ED 275 771 TM 860 693

AUTHOR Katzenmeyer, Conrad G.
TITLE The Federal Role in Encouraging State-by-State Achievement Comparisons.
INSTITUTION Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Apr 86
NOTE 8p.; Paper presented at the Annual Meeting of the National Council for Measurement in Education (San Francisco, CA, April 17-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Tests; College Entrance Examinations; *Comparative Testing; Data Collection; *Educational Policy; Educational Research; Elementary Secondary Education; Federal Government; *Federal Programs; National Surveys; *Research Design
IDENTIFIERS American College Testing Program; *Indicators; *National Assessment of Educational Progress; Scholastic Aptitude Test

ABSTRACT

With the growing interest in comparing student achievement state by state, there is a need to develop better techniques for making such comparisons. Approaches available now, particularly the use of Scholastic Aptitude Test (SAT) and American College Test (ACT) scores, have obvious shortcomings but nonetheless continue to be prominent because there is nothing more defensible. In this paper, the role of the federal government in encouraging new approaches is discussed, particularly the exploration of the use of state assessments in cross-state comparisons, and the extension of National Assessment of Educational Progress (NAEP) to a nationally representative sample for this purpose. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275771

The Federal Role in Encouraging State-by-State Achievement Comparisons

by

Conrad G. Katzenmeyer
Office of Research
Office of Educational Research and Improvement
U.S. Department of Education

Abstract

With the growing interest in comparing student achievement state-by-state, there is a need to develop better techaniques for making such comparisons. Approaches available now, particularly the use of SAT and ACT scores, have obvious shortcomings but none-the-less continue to be prominent because there is nothing more defensible. In this paper, the role of the federal government in encouraging new approaches is discussed, particularly the exploration of the use of state assessments in cross-state comparisons, and the extension of NAEP to a nationally representative sample for this purpose.

Paper presented at the National Council for Measurement in Education meetings, San Francisco, California, April, 1986

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Conrad Katzenmeyer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 860 693

The Federal Role in Encouraging State-by-State Achievement Comparisons

by
Conrad G. Katzenmeyer
Office of Research
Office of Educational Research and Improvement
U.S. Department of Education

State-by-state comparisons appear inescapable, no matter the cautions voiced about the interpetations to be made. The Reno Gazette-Journal of March 3, 1986 published under the headline "Nevada at bottom of education rankings" a recounting of the latest findings from the Secretary's wall chart, with a comment from the Chief State School Officer that the rankings are "probably not a good indicator of (Nevada's) overall...school program" but should not be dismissed. On the other hand, the Milwaukee Journal reported ecstatic comments about Wisconsin coming in first in the ACT rankings.

While competition and comparisons have long been a part of our culture, state-by-state comparisons of educational achievement have only recently been made. The Council of Chief State School Officers traditionally opposed such comparisons for student achievement, and The National Assessment of Educational Progress was designed as it was to make such comparisons impossible.

The purpose of this paper is to review the role of the federal government, specifically the Department of Education, in the current discussion of cross-state comparisons. Also, I will attempt to give some indication of the directions the Department of Education might take in the near future. However, I must include two caveats. First, the Department is now in the process of formulating a long-term policy regarding state comparisons, which will be part of a much larger strategy for data collection. Very little can be said for certain at this time. Second, I am not directly involved in the current discussions. Since the reorganization of OERI relocated all of the regular data collection programs under the Center for Statistics, the Office of Research where I am located has had no direct connection with these activities. While this may serve to make me a more objective reporter, it certainly does not help to make me a better informed one.

History of Federal Involvement

The U.S. government has always had responsibility for collecting information about education throughout the country. In fact, this is the primary purpose for a federal role in a system that is under state and local control. The U.S. Dapartment of Education regularly collects statistical data through the Center for Educational Statistics in OERI (previously the National Center for Educational Statistics), much of it from state departments of education, some directly by various contractors. The results are routinely reported through a variety of mechanisms. While technical issues are raised at times about the quality of these data, there is generally little contrvercy about their purpose or use.

This pattern of tranquility and lack of confrontation changed with the publication of the "wall-chart" by Secretary Bell in 1983, a compliation of state level statistics that ranked the states' performance. This was immediately criticized by state officials as well as statisticians on the grounds that misinterpretations were likely to made, particularly in regard to the student achievement outcomes. Lacking other state level measures, the wall-chart contained SAT and ACT scores. Because these tests are vountarily taken, and the percentage of students differs so much across states, no direct interpretation of the results is possible. Participation rates differed within states between 35 and 86% for the ACT and 33 and 63% for the SAT. Since it gives its own college entrance test, The state of Washington failed to have 20% of its students take either and thus was not included in the academic achievement rankings. The Department recognized the limitations in these statistics, but because there are no more better alternatives, the ACT and SAT rankings have continued to be included in subsequent compilations of the wall chart.

At approximately the same time, NCES launched an indicators project. The intent was to identify a limited number of key measures of school status that could regularly be collected and reported. Although there has bound to be overlap with statistics already being collected, the purpose was different. Instead of just a descriptive summary of educational status, analogous to the census, educational indicators were to be more like economic indicators, providing trend data and a picture of the system as a whole. Greater emphasis was also placed on relationships between the measures.They were grouped according to outcomes, resources, and context of education when an initial set of indicators were identified.

For the critical achievement outcome measures, however, the indicators project had some of the same difficulties as the wall chart. Although it drew on both the NAEP data and the international comparison studies, the Indicators project was forced to use SAT and ACT scores for state comparisons. Thus some of the most important policy implications of the project depended on these badly flawed indicators.

One way to solve the problem of achievement indicators would be to expand the current NAEP testing program to draw a respresentative sample of students in each state. The contractor ETS, and its advisory board, are quite willing to do this, but the cost would be much higher than the current contract. Also, the intermittent schedule of testing that NAEP uses makes it less desireable as an indicator.

Recognizing these difficulties, and the fact that a substantial amount of achievement testing is being done by state and local school districts, NCES commissioned the Center for the Study of Evaluation to do a study on the feasibility of using state data as national indicators of educational quality. Analyzing whatever documentation could be collected from the states, whether actual tests, test specifications, or other information, the CSE study group concluded that there appears to be enough common content among state tests to make comparisons possible. It would be feasible to develop a common core of items, scaled by IRT methods, that could serve as a linking mechanism. However, this could only be done through an evolutionary process, as the differences in procedures across states would make it impossible for even most of the states to immediately participate in such an effort, no matter how conceived.

This brings us to the Fall of 1985, which coincided with the reorganization of OERI and the consolidation of all statistical data collection under the Center for Educational statistics. Three other events, should also be noted.

o Secretary Bennett and Assistant Secretary Finn called for a substantial increase in the NAEP testing program-more frequent testing, particularly in the basic skills; testing of adults; and increased subject matter testing.

o The Council for Chief State School Officers reversed its position against state-by-state achievement comparisons. In order to devise means for implementing this change, CCSSO established a Center on Assessment and had a committee studying the ways that cross-state achievement comparisons might be made.

o Three southern states completed a mini-NAEP study, where a shortened version of the NAEP reading assessment was administered by these states, at their expense, as a feasibility study. This is to be repeated in 1986 with a NAEP writing assessment.

Current Plans

Consolidation of the statistical data collections has permitted CES to look comprehensively at what that total data set is, and how it might be altered to meet needs more effectively. CES has been exploring this since early 1985, recently publishing a Working Paper on a Plan for the Redesign of the Elementary and Secondary Data Collection Program(March 27, 1986). A number of papers had been commisssioned to analyze the existing data set. Authors of these papers concluded that the data available from CES are deficient in the following respects:

o Comprehensiveness: there are notable gaps in coverage, particularly for student achievement, teacher data, and school finance. With respect to NAEP, the gaps particularly noted were the infrequency of testing for most academic subjects and the lack of state representative samples. CES also collects achievement information in the National Educational Longitudinal Study(NELS) and the High School and Beyond study but achievement is not a central component of either and the content coverage is quite limited.

o Integration- The data collection efforts are generally conducted independently, making it impossible to establish interrelationships.

o Inaccuracy and Noncomparability-Errors are notable in the data set due, in part, to the various agencies involved in data collection with differing definitions and procedures.

o Representativeness- much of the data can not be disaggregated below a national level. This is particularly true in regard to achievement and other student characteristics.

o Timeliness-There have been significant delays in collecting, compiling and summarizing data, often leading to the publication of separate indicators from several different years in the same reports.

In response to these problems, CES has proposed a redesign of the Elementary/Secondary Integrated Data System. This would include a sample of 4200 public and private schools with stratification at the 4th, 8th and 12th grades. The school would become the critical sampling unit, with samples of students, teachers, parents and administrators drawn from these schools. About 25% of the sample schools would be replaced annually.

In regard to achievement testing, a revised NAEP program would be established, including shortened versions of the tests. Reading and mathematics would be alternated yearly, with a second academic subject assessed each year. These instruments would yield a single score in the particular subject matter. The intent

is to keep assessment time in each area to an hour or less, but still be able to continue the time trend analysis from previous assessments.

A major strength of the new design is its greater capacity for providing relationships. Since students, teachers and parents would all be sampled from the same schools, it should be possible to directly analyze interrelationships. While the sample would include individuals representing all greade levels, there would be special concentration on the target grades.

With cluster sampling of schools, it should also be possible to collect, summarize and report the data in a timely manner. CES would be responsible for collecting the school level data.

Finally, the redesign calls for a good deal of background information on students, teachers, schools and programs. This continues a trend that was started some years ago with NAEP, but represents a departure from the previous NCES practices.

However, there are also a number of questions and concerns that will need to be addressed.

1. Is it possible to create a shortened or streamlined NAEP test for the areas desired? In order to relate back to previous administrations of the NAEP instruments, the tests will need to be equated, most likely through IRT methods. A shortened reading test has already been successfully scaled, but this is yet to be done in mathematics. Whether it is possible in other areas, particularly in science, is unknown.

2. What happens to the current NAEP assessment? There is an implication to be drawn that NAEP as it is now done would be phased out, but this is not certain. It appears clear that item development would be continued for the new instruments, regardless of what happens to the current NAEP. It could be that a version of NAEP similar to what is now available would be continued for states that do not wish to build their own assessment program. How much financial responsibility for this the federal government would assume is not known.

3. Can state comparisons be made? At the outset, there will be no plans to make such comparisons in student achievement. State comparisons would be possible for other indicators that are not sampled at the school level. Initial samples would be drawn in all states, but the schools would not be state representative, except perhaps in a few of the larger states. Achievement comparisons would eventually come through an evolutionary process, hopefully with considerable assistance from the states.

If states are to become involved in the data collection, then there will need to be some new procedures developed. The trade-off of quality versus cost will need to be explored. It might be possible for the federal government to do some of the data collection for its purposes, and then provide the instruments to the states for their purposes.

4. How useful will the new achievement test be to the states? There is certainly a strong hope that the states will play a major role in this effort; the tests are to be designed so they could readily be incorporated into the states' assessment programs. States should be able to scale their instruments to the national test and thus give more fine-grained breakdowns within their state while providing the data for comparisons across states on the national standard. However, this will be possible only for the abbreviated tests. Just how much enthusiasm, and resources, the states will have for comparisons based on the streamlined measures is open to question.

I draw two conclusions from the recent federal proposal. First, there appears to be a shift in the NAEP Assessment Program, making it a federal rather than a national assessment. Revisons suggested in NAEP will neet federal needs more adequately than is now the case. Whether these revisions will meet others` needs as adequately is not as clear.

Second, it does appear clear that the states will have to take a more active role in making cross-state comparisons that are meaningful for them. That this is already occurring is an indication that the federal government may not need to be a central force in actually creating such comparisons.